

周报

2015-09-20

本周总结:

1 网易项目:

算法:

算法方面因为之前算法存在了一些问题，不太能适用，因此换个方向对方法进行优化：1) 尝试使用自适应k值的聚类方法 2) 自适应类特征的权重。

自适应特征值的方法是源于论文review里提出的问题，即k值如何确定的问题，之前使用过层次聚类的方法，然而层次聚类从本质上讲依然是一种划分的聚类方法，只是一层一层划分下来之后聚类的个数每一层不同，可以选择不同的层次从而确定聚类个数，然而其实本质上来说还是要制定聚类个数的。而另一些算法，比如：DBSCAN谱聚类算法，则不需要确定的一个k的个数，而是通过自适应的方式通过点特征或者分布密度计算得出具体类别的信息。

另外之前我们进行计算的时候是将所有的属性都参与聚类，但是其实缺少了分析的过程，并不是所有的属性都值得拿来进行聚类的。所以在聚类的方法上，我们也可以考虑一些能够在聚类时自适应不同属性所占权重的聚类方法。

代码:

代码方面，本周开始计划重构代码重定需求：

1) 筛选功能，筛选功能希望能够加入更多的可视化元素，之前我们进行筛选的时候，是通过设定值来进行筛选的，这样功能和交互太过单一，也缺少可视化元素。我希望能在此功能上加入一些可视化的图表，展示出当前属性的分布情况，同时通过在分布图上进行交互来达到筛选的目的。

2) 聚类属性选取，这一点是巫老师之前提出过的，他希望能在聚类的时候能够进行部分的聚类，通过用户选取特定属性，从而更好展示用户关注的特征。

3) 数据解析上的优化。之前在做数据解析的时候，代码写的可移植性十分的差，同时载入速度过慢，也缺少缓存的机制，因此，整个计算还是比较慢的，因此希望能重构代码，数据解析更加清楚，可移植性更好。

4) 聚类方法上，因为我们希望加入实时的部分聚类功能，因此算法是一定需要进行优化的，不能使用传统的单线程基础版本的算法，所以可能要对确定的算法进行优化，让他能更快的跑好我们的任务。

2 网易项目：

调研：

1) openstreetmap：对openstreetmap进行调研，openstreetmap与谷歌地图、百度地图、高德地图不同之处在于其开源性与可编辑性，OpenStreetMap是一款提供完全免费地理数据且没有法律与技术约束的全民地图，用户可以想wiki百科一样对地图进行编辑，也可以随意下载地图的离线数据进行使用，在网站上搜了一下，一个实时的离线地图的数据包大概在80g左右。对于不同的应用和不同的需求，离线数据包大概在30-45g左右，如果用户想要进行地图更新，可以直接下载更新包，对之前数据进行更新

2) 后端搭建计划：考虑到我们之后的数据规模和需求，后端打算使用java语言，利用spring框架进行搭建，同时使用maven对项目进行管理。这主要是因为考虑到java的扩展性更强，尤其当要做大规模数据的时候，一些分布式存储系统，分布式计算系统对java都有很完备的接口支持，同时在处理复杂逻辑的时候java也有更好的表现。